The gene encoding the sperm-specific basic nuclear protein ϕ_o from sea cucumber

Eva PRATS & Luis CORNUDELLA *

Department of Molecular and Cell Biology,
Centro de Investigación y Desarrollo del CSIC, Jordi Girona 18-26, Barcelona, E-08034, Spain
* correspondence and reprints

ABSTRACT

The gene encoding the sperm-specific protein ϕ_o from the sea cucumber *Holothuria tubulosa* has been cloned and characterized. Sea cucumber sperm chromatin displays a somatic-like histone complement and is beaded with a constant 227 bp DNA linker length, second to the longest repeat of sea urchin chromatin. Protein ϕ_o, a small basic protein reminiscent of the C-terminal tail of histone H1, appears at the onset of spermiogenesis and accumulates in ripe sperm. The ϕ_o gene displays a coding frame interrupted by three large intervening sequences which combine to make it the longest for a sperm-specific protein yet reported (*ca.* 17.7 kb). The identified gene is present as a single copy and, in turn, encodes a polyadenylated transcript. The protein ϕ_o specified by this gene has 77 residues, reproducing unaltered the partial amino acid sequence of the protein previously determined. The structural arrangement and content of the ϕ_o gene are totally unrelated to cell-cycle regulated histone gene structure. Instead, it combines several features common to replication-independent genes coding for histone variants and even to protamine genes. Inference is made about the potential implications of this divergence in gene arrangement as regards chromatin transitions and modulation of gene activity occurring during spermiogenesis.

RÉSUMÉ

Le gène codant pour la protéine basique nucléaire ϕ_o spécifique des spermatozoïdes chez l'holothurie

Le gène codant pour la protéine basique nucléaire ϕ_o spécifique des spermatozoïdes chez l'holothurie *Holothuria tubulosa* a été cloné et caractérisé. La chromatine du spermatozoïde d'holothurie présente un complément d'histone de type somatique et forme un chapelet avec un intervalle constant de 227 paires de bases d'ADN, qui est la seconde plus longue répétition de chromatine d'échinoderme. La protéine ϕ_o, une petite protéine basique qui rappelle l'extrémité C-terminale de l'histone H1, apparaît au début de la spermiogenèse et s'accumule dans les spermatozoïdes mûrs. Le gène ϕ_o comprend une région codante interrompue par trois grandes séquences intercalaires qui se combinent pour en faire le gène le plus long connu actuellement pour une protéine spécifique des spermatozoïdes (environ 17.7 kb). Le gène identifié est présent en copie unique et code pour un transcrit polyadénylé. La protéine ϕ_o spécifiée par ce gène a 77 résidus parmi lesquels on retrouve sans erreurs la séquence partielle d'acides aminés précédemment déterminée pour la protéine. L'organisation structurale et le contenu du gène ϕ_o sont sans relations avec la structure des gènes des histones régulés par le cycle cellulaire. Au contraire, ce gène combine plusieurs caractéristiques communes aux gènes indépendants de la réplication codant pour des variants d'histones et même aux gènes de protamines. La conséquence potentielle de cette divergence dans l'organisation des gènes en ce qui concerne les transitions de la chromatine et la modulation de l'activité du gène pendant la spermiogenèse est discutée.

PRATS, E. & CORNUDELLA, L., 1995. — The gene encoding the sperm-specific basic nuclear protein ϕ_o from sea cucumber. *In*: JAMIESON, B. G. M., AUSIO, J., & JUSTINE, J.-L. (eds), Advances in Spermatozoal Phylogeny and Taxonomy. *Mém. Mus. natn. Hist. nat.*, **166** : 491-500. Paris ISBN : 2-85653-225-X.

Eukaryotic chromatin is a macromolecular nucleoprotein assembly essentially composed of DNA and basic proteins. Histones, which are relatively conserved through evolution, are the genuine protein components in chromatin of non-proliferating cells and are involved in the organization of the chromatin fibre into various structural hierarchies (for full review see [42]). In stark contrast, DNA in male germ cell lineages appears to be bound by widely diverse basic proteins as regards chemical composition and number, giving rise to a seemingly striking variety of protein molecules along the different zoological groups [32], spanning from those species which retain histones that are close to the somatic types, to those in which they are fully displaced by more basic proteins like nucleoprotamines [7]. The chemical data reported comprise a variety of vertebrate species, from fishes [9, 17] to mammals [25]. It is apparent that an enhanced basicity of the sperm proteins favours a tighter packaging of DNA. This requirement in spermatozoa seems obvious as a means of protecting the genomic complement. Likewise, it may also be an evolutionary adaptation of sperm to endure long-term storage and transport in the absence of DNA repair mechanisms. However, the functional significance of basic protein diversity in sperm and its effect at the molecular level are still ill-defined. DNA packing may not be the exclusive role. The actual existence of germ-line variants and sperm-specific protein types argues for more discriminating assignments such as fine-structural transitions of chromatin related to modulation of gene activity and its final quiescence during spermiogenesis [14].

The variability of sperm nuclear proteins is of unknown origin. A hypothetical evolution of some of these proteins from histone H1 has been put forward on the basis of the protein composition of the sperm of some bivalve molluscs [4]. Histone H1 and its many subtypes constitute the most heterogeneous of the histone classes [41]. Its largest evolutionary sequence-variation appears confined to both the N-terminal and C-terminal extensions of the molecules whereas the hydrophobic central globular core remains fairly well conserved [13]. This asymmetric organization has led to the suggestion that the carboxyl-terminus is involved in the higher order compaction of chromatin [1]. Models to test that assumption are provided by marine invertebrates, particularly echinoderms and molluscs. These organisms deserve particular mention since their somatic histones apparently coexist with both sperm-specific variants and protamine-like molecules, different from fish or mammalian protamines [3, 32, 43]. These protein molecules mostly fit into classes 3 and 4 of BLOCH's cytochemical categorization of nucleoproteins in mature sperm cells [7]. These two types have been ulteriorly combined into a rather heterogeneous group of intermediary sperm-specific types, plausibly representing transitions from histones to protamine-like molecules [32].

Although the prevalent cellular histones are encoded by a highly reiterated multigene family whose expression is tightly coupled to DNA replication, histone-variant genes tend to be present in single or few dispersed copies not subjected to cell cycle regulation [11, 23]. Already classification schemes based on regulatory correlations have been devised [43]. Nonetheless, very little is known about the organization of tissue-specific, variant-histone genes, along with the evolutionary origin of nucleoprotamine and protamine-like genes [30].

It is important to address these questions and obtain new evidence concerning histone to protamine transitions and the genes that encode them, aiming to understand at the molecular level their differential function and its influence on the structural organization of sperm chromatin. Our work has been involved in the analysis of chromatin from the germinal tissue of the echinoderm *Holothuria tubulosa*. During sperm maturation there is no bulk replacement of the histone complement, transitions being restricted to the addition of a sperm-specific, arginine-rich H1 variant [31] and the presence in ripe sperm of a small basic protein termed ϕ_0 [40]. The latter has an amino acid composition reminiscent of the carboxy-terminal region of sea urchin H1-S, provided that Arg is considered equivalent to Lys [2]. Incorporation of protein ϕ_0 into chromatin occurs in the terminal stages of spermiogenesis [10], representing about 4% of the histone moiety of the mature spermatozoa. Nucleosome organization remains invariable throughout sea cucumber spermatogenesis with a constant DNA linker length of 227 bp [15] consistent with sea urchin

nucleosomal repeats, which exhibit the longest lengths ever measured. The isolation and sequence determination of a cDNA for *H. tubulosa* protein ϕ_o have been previously reported [33]. In the present paper we describe the molecular cloning and characterization of the gene encoding this protein specific to the sea cucumber sperm chromatin. This is the first gene coding for a histone-to-protamine transition protein to be identified.

MATERIAL AND METHODS

Living organisms. Male specimens of the sea cucumber *Holothuria tubulosa* were collected periodically off the catalonian shore during the breeding season, moved live to the laboratory in cold seawater and held in 8°C seawater until used. Excision of gonads and sperm collection were performed as detailed elsewhere [36].

Isolation and purification of genomic DNA. High molecular weight genomic DNA was extracted from fresh sperm suspensions essentially as described [15]. Briefly, suspensions were treated with proteinase K (50 µg/ml) overnight at 37°C. After incubation, samples were deproteinized by successive phenol and chloroform extractions and the aqueous phases precipitated with ethanol. The DNA was further purified by Cesium chloride banding and subsequent dialysis.

Construction and screening of a sea cucumber genomic library. For construction of the *H. tubulosa* genomal library, purified sperm DNA was subjected to a partial digestion with *Mbo*I to generate fragments with *Bam*HI-compatible overhangs and subsequently size-fractionated by sucrose gradient centrifugation. DNA fragments in the size range 12-20 kb were pooled and ligated to dephosphorylated lambda-based Charon 35 (λCh35) replacement vector [26] linearized with *Bam*HI. Ligation reactions were carried out at a vector to insert molar ratio of 2:1. Recombinant phages were encapsidated and used to transform *E. coli* 555 *recA*$^-$ cells yielding a titre of $3x10^5$ plaque-forming-units (pfu) per µg of ligated DNA. The genomal library was screened by *in situ* plaque hybridization [5]. Plaques at a density of 10^4 pfu were replicated onto nitrocellulose membranes and screened with a 441 bp long *H. tubulosa* ϕ_o-cDNA clone [33] labelled by random-priming with the Klenow enzyme [21].

Positive plaques were purified by plating at decreasing densities and the isolated phages were grown by cascade infection and banded onto ethidium bromide-containing CsCl gradients. The resulting DNA was purified by phenol and chloroform extractions and used for further analysis. All recombinant DNA manipulations were carried out by standard procedures [37] and conducted in accordance with established guidelines for recombinant DNA research.

Restriction analysis and Southern transfers. DNA from positive recombinant clones was digested with selected endonucleases and pairwise combinations thereof. Where required, restriction fragments were electrophoresed on agarose gels, transferred to nylon membranes by alkaline blotting [34] after partial depurination and screened with six different regions of the ϕ_o-cDNA cloned insert as probes.

ϕ_o gene number. To assess the copy number of the ϕ_o gene in the haploid sperm genome of *H. tubulosa*, genomic DNA was digested, independently or in combination, with various endonucleases not cleaving inside the ϕ_o-cDNA sequence. Restriction fragments were subsequently electrophoresed on 0.5% agarose gels, blotted onto a nylon membrane, and hybridized to labelled probes prepared from the ϕ_o-cDNA clone. The ϕ_o gene number was derived by comparison of the number and intensity of the autoradiographic signals measured by densitometric analysis, with those from graded amounts of the cloned ϕ_o-cDNA equivalent to integer-copies per haploid genome of the cDNA sequence. Standards of cDNA were electrophoresed in parallel, supplemented with a mass excess of sheared calf thymus DNA to compensate for the amount of restricted sea cucumber sperm DNA loaded on each gel slot.

Plasmid subcloning and nucleotide sequence analysis. Genomic DNA restriction fragments of appropriate size identified by hybridization analysis as containing ϕ_o-cDNA sequence tracts were excised off the gel, purified further on low-melt agarose, and ligated into the phagemid vector Bluescript $^+$SK. Chimeric recombinants were used to transform competent *E. coli* XL1-blue *recA*$^-$ cells and were selected as AmpRTcR:Lac$^-$ phenotypes. DNA inserts from the recombinant plasmids were sequenced by the dideoxy chain-termination procedure [38] using the Sequenase system with forward and reverse primers for both orientations. Computer analysis was performed using the MicroGenie sequence analysis software (Beckman, USA).

RESULTS AND DISCUSSION

Spermatogenesis in the sea cucumber *H. tubulosa* is a rather simple process. Chromatin from ripe spermatozoa retains the five somatic-type histones in normal relative amounts accompanied by the highly basic protein ϕ_o (average M_r = 8640), structurally related to histone H1 [2]. We have previously reported the characterization of a clone carrying a full length ϕ_o transcript, isolated from a cDNA expression library made from the poly (A$^+$) fraction of total RNA extracted from immature gonadal tissue and screened with polyclonal anti-ϕ_o antibodies [33]. The 441 bp cloned cDNA encompassed a continuous open reading frame for a basic polypeptide of 77 residues whose sequence conformed to the partial amino acid sequence of ϕ_o previously established. Likewise, poly (A$^+$) selected RNA yielded a product electrophoretically

comigrating with protein ϕ_O upon *in vitro* translation in wheat germ, cell-free extracts, whereas Northern blot analysis detected only a 0.6 kb ϕ_O-mRNA transcript homologous to the cDNA probe.

Isolation of the sea cucumber ϕ_O gene

The genomal library of *H. tubulosa* sperm DNA cloned into the λCh35 vector was screened by hybridization with the ϕ_O-cDNA cloned insert. The screening of about 250 000 plaques yielded four positive clones with inserts of ~16.5, 14.4, 13 and 14.5 kb, named λHt7, λHt2, λHt1 and λHt8, respectively. The four recombinants were next subjected to endonuclease restriction and Southern blot analysis using six cDNA-derived probes encompassing specific regions of the ϕ_O-cDNA clone: (i) the two asymmetric segments resulting from the cleavage along the *Eco*RI site internal to the ϕ_O-cDNA insert, namely, the 248 bp leader (*5'* probe) and the 193 bp trailer (*3'* probe) fragments, encoding the first 69 and last 8 amino acids of protein ϕ_O, respectively; (ii) the 81 bp long *Eco*RI-*Pst*I fragment comprising the 5'-flanking region and the 42 bp sequence coding for the 14 amino acids inclusive of the initial methionine residue heading the N-terminus (*P* probe); (iii) the *Dde*I-*Eco*RI fragment of 81 bp representing amino acids 43 to 69 (*D* probe); (iv) the two segments of 113 bp and 80 bp in length (*X1* and *X2* probes) resulting from the nibbling of the *3'*-probe at the single *Xba*I site. Probe *X1* contained the sequence for the last 8 amino acids of the C-terminus plus 88 bp of the adjacent downstream extension whereas probe *X2* consisted of the final 80 bp of the 3'-noncoding region of the ϕ_O-cDNA clone. Those restriction fragments shown to carry ϕ_O sequences were subsequently subcloned for further DNA sequence analysis.

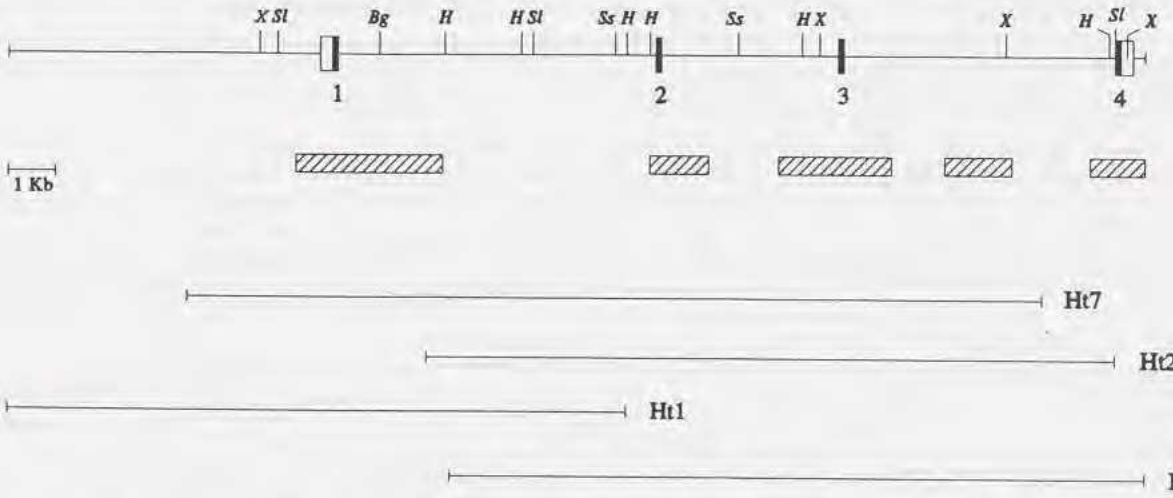

FIG. 1. — Isolation and characterization of the *Holothuria tubulosa* ϕ_O gene. The restriction endonuclease map and organization of the ϕ_O gene are shown. The four positive isolates picked out from the sperm DNA library of 10-20 kb *Mbo*I partials cloned into the *Bam*HI replacement vector λCh35, were digested with several endonucleases and combinations thereof. Restriction fragments carrying ϕ_O-cDNA sequences were identified by hybridization. Positively reacting fragments were further purified, subcloned into the phagemid Bluescript $^+$SK and sequenced by the dideoxy chain-termination procedure of Sanger. Filled boxes indicate the relative positions of exons encoding ϕ_O sequences consecutively numbered 1 to 4, and open boxes the 5' and 3' flanking regions homologous to the noncoding extensions of the cloned ϕ_O-cDNA. The hatched rectangular boxes highlight those regions of the gene that were completely sequenced either on both DNA strands or repeatedly in one direction. *B, D, E, H, P, Sl, Ss*, and *X* denote *Bgl*I, *Dde*I, *Eco*RI, *Hin*dIII, *Pst*I, *Sal*I, *Sst*I, and *Xba*I restriction sites, respectively. The thin lines depict the positions of the four strongly hybridizing genomic clones (λHt 1 - 8) used to map the ϕ_O gene.

The results of the restriction mapping are shown in Fig. 1 and can be summarized as follows. The entire ϕ_O gene appeared split into four distinct exon sequences scattered along the lengths of the clones. The latter displayed partial overlaps differing in extension. λHt7 harboured the 5'-proximal exons 1, 2 and 3. The central exons 2 and 3 were also present in λHt2 and λHt8. In addition, the former contained the first 79 nucleotides of the fourth exon whereas the latter spanned its entire coding sequence extending 831 nucleotides beyond the stop codon. The shortest clone λHt1 carried only exon 1 centrally positioned within the genomic DNA insert.

On the basis of these results it was feasible to correlate the four recombinants and to conclude that the coding sequence of the sea cucumber ϕ_O gene is interrupted by three long intervening sequences which amount to 16.2 kb in total length (see Fig. 1). These unusually large introns combine to make the ϕ_O gene the longest for a sperm-specific protein (*ca.* 17.7 kb) so far reported. The overall organization of this sea cucumber gene appears to significantly diverge from that of the intron-less histone genes [27] and also from the arrangement of mammalian, single-intron protamine genes [30].

Sequence analysis of the ϕ_O gene

The nucleotide sequence of the sea cucumber ϕ_O gene is shown in Fig. 2. The coding sequence is discontinuous and encompasses an open reading frame for a basic protein of 77 residues, interrupted by three introns involving canonical splice junctions [24], specifically those assigned to invertebrates [22]. The first and second introns of 6.8 and 4 kb in respective length, are inserted within codons 9 and 15, respectively. The third intron is 5.4 kb long and is positioned contiguous to codon 41. The complete coding sequence of the ϕ_O gene is identical to that of the cDNA clone previously reported [33] but for six nucleotide substitutions (97.9% homology). Five changes are conservative since they correspond to third-base degenerates of the most common synonym triplets, involving C for T and A for G conventional exchanges with no ensuing alteration of the assigned amino acid. The only relevant nucleotide substitution occurs at codon 61 and involves a G for A replacement in the first base causing a change of amino acid assignment. The overall level of sequence conservation observed argues for a stable organization of this gene.

The deduced primary structure of the encoded protein corresponds exactly with the ϕ_O sequence specified by the cDNA clone, exclusive of the noted single alteration at codon 61 (98.7% homology). The cDNA sequence contains the triplet GCC for alanine in this position while the ACC counterpart in the gene codes for threonine. Most likely this difference arises from the DNA polymorphism detected in echinoderms [12] which is reflected in the well-documented intraspecific microheterogeneity found in a substantial variety of sperm variant proteins from marine invertebrates [28] as well as in the protamines of trout [18].

The close sequence homologies between the ϕ_O-cDNA and the cloned gene are endorsed by the identity of the deduced protein sequences which, in addition, reproduce wholly unaltered the partial amino acid sequence of protein ϕ_O previously established. The coincident similarities observed sustain the conclusion that, in actuality, the cloned gene encodes the sperm-specific protein ϕ_O in *H. tubulosa.*

Comparison of the nucleotide sequences in the noncoding regions of the ϕ_O gene with those of replication-dependent histone genes reveals that the former lacks the conserved motifs defining the S-phase regulated histone gene structure such as the downstream hairpin loop sequence at its 3' proximal purine-rich tract required for 3' processing of the histone transcripts [6]. Instead, the leader and trailer regions surrounding the ϕ_O gene combine several structural elements found in both replication-independent histone variant as well as protamine genes (see Fig. 2).

The region upstream of the initiation codon contains an atypical TATA motif identical to the TTCAAA box identified in the cell-cycle independent $H2A_F$ histone gene that codes for an extreme H2A variant in chicken, whose transcript is polyadenylated [16]. Significantly, both

elements are found similarly positioned 144 nucleotides upstream from the initiator triplet. There is another potential TATA motif with perfect homology to the non-canonical TTAAAT element present in both the chicken and duck H5 genes [19, 35]. This sequence starts at position -205, 65 and 28 nucleotides further upstream than the mentioned homologues, respectively. Another general feature required for promotion of transcription by RNA polymerase II is a CAAT sequence often located between 70 to 90 bp upstream of TATA sites [8]. In this regard, the leader region of the ϕ_O gene displays a potential CAAT motif (-221 to -218) located 73 nucleotides upstream relative to the $H2A_F$-like TTCAAA box. Another motif shared with leader regions in protamine genes is the TGACGTCA sequence found far upstream in the ϕ_O gene (-489 to -482). This *cis*-acting element, usually referred to as a cAMP regulatory element (CRE), is strictly conserved in all protamine genes and it is considered essential for the biological activity of cAMP-regulated enhancers [30]. Since spermatogenesis in echinoderms is known to be under hormone control probably involving cAMP [39], such a regulatory signal might well represent a link between hormonal signals and the expression of sperm-specific genes such that of protein ϕ_O.

The downstream extension of the ϕ_O coding sequence is devoid of the highly conserved structural features of the replication-dependent histone genes required for the 3' end formation of histone transcripts. Instead, similarities with the equivalent regions in protamine genes are encountered. First, three potential polyadenylation signals are present, starting at positions 306, 392 and 396 respectively, 3' to the TAA stop codon. The last two elements consist of the heptamer AAATAAA which appears repeated with a trinucleotide overlap. This heptameric sequence motif bears a perfect homology with the conserved polyadenylation signal found in the protamine genes from salmon and trout [29]. Nonetheless, the significance of the close similarities encountered between the organization of the ϕ_O gene and that of genes coding for extreme histone variants or even for protamines remains to be unambiguously defined.

Genomic content of the ϕ_o gene

The copy number of the ϕ_O gene was determined by Southern blot hybridization analysis of sperm DNA restriction digests with the 81 bp long *Eco*RI-*Pst*I fragment (probe *P*) of the ϕ_O-cDNA clone comprising the 5'-flanking extension and the initial 42 bp of the coding sequence, labelled by random priming. A set of endonucleases lacking cleavage sites within the cDNA sequence was selected for the single and double enzyme digestions. DNA restriction fragments were electrophoretically resolved in conjunction with varying amounts of the cloned ϕ_O-cDNA insert, diluted with a mass excess of sheared heterologous DNA to make up for those of restricted sperm DNA loaded on the gel lanes. The amounts of the cDNA standards, equivalent to one and four copies of the respective sequence per haploid genome, were inferred from the DNA content of the haploid genome (C-value = 3 x 10^9 bp) of *H. tubulosa* sperm previously determined [36].

Hybridization patterns from both single and combined enzyme restrictions yielded in every case only one size class of DNA fragment positively hybridizing with the cDNA probe (Fig. 3).

FIG. 2. — DNA sequence of the sea cucumber ϕ_O gene and flanking regions. The 5'→3' nucleotide sequence of the non-transcribed strand (i.e. mRNA-like) is given along with the amino acid sequence derived from the ϕ_O coding region, shown above the nucleotide sequence. The abbreviation *ini* and the asterisk mark the respective positions of the initiator and the stop codons. The noncoding leader and trailer extensions are numbered with negative and positive numerals begininng at nucleotides 5' adjacent to the initiation codon and 3' proximal to the stop triplet, respectively. Coding triplets are denoted with numbers in *italics*. The slash symbols mark the normal donor splice site junctions. Positions of the putative CRE, CAAT and TATA elements as well as the poly(A) addition signal discussed in the text, are doubly underscored. Most of the extensive intron sequences have been removed for clarity.

Exon 1

```
                                -700                                                    -650
GACTGCTCACGAGTGATAGCGCCCAAACAAGCGTCCAAGAGGCCGAACAGACGCTAAGGCTCGCTACGGCGCGTTAGCAT
                                                        -600
GTACAAAATTGCCTCGGCGTTTTCACTAGTATACGCGTTTCTACGGCGATGGAATCGACTTGAGTAAGCTGACCTACATC
                    -550                                                    -500
ACGTGACCTTCCATGATGGCTTCACAACAACACTGTGGGATGCGCCTTCCTGTTATATAGTTCAGTGGAGCCGATGCTTA
                                        -450
TGACGTCATAATTTGGTGTAGAAGTCTAGTCCACTGCATATTTCCAAACGTGAAGAAACGGTTTGAGACCGAATCAAATG
      -400                                                              -350
CATTTTTCTTACCAGCTTACCGCGCTCTGGAATACGATGAGTATGTATGCAATCCTCAGTTCAAAGCAAAGTGGACGCAC
                                -300                                                    -250
TACGTTTGGACACTGTACCTAGCCTACATTATTCATGTTCTTGTTCCCTTTATGCACGTAAGGACGCAACCCGATGCGAC
                                                        -200
GTTTCTGAAACGGGCACTTAGACGCGCGCAATTTGTAATCACGTTTAAATCATAGTATTTGGTGCGCGTAGCTATAGGTG
                    -150                                                    -100
CGTTTATGCGCCTCACTTTGTAAAATTCAAAAAAATAACAAATATTGTTTTCAGTTTTTAAAGAACCGTTACTTGATCGC
                                        -50
TTACAGCCGAGCAACTAAGACGTTGGTCCTACGCACTGCCCAGTTTTGATTCCCCCTTGTGTCGGAAATTCCAACTACAA
      -1                    5
TCAATAATC ATG GTA GCC AGA CGA CAA ACA AAG AAA G/GTAAATAAAGGGAACAATATATGCAAGGCGTT
          ini Val Ala Arg Arg Gln Thr Lys Lys A
```

Exon 2

```
                                    10                  15
TTTTATTTCTTCTTTCTCACTTCACAG CT AGG AAG CCT GCA GCC AG/GTGAGTGAATACAATTTAAATTTTAT
                               la Arg Lys Pro Ala Ala Ar
```

Exon 3

```
                                                                20                  25
TAACTCAAGACTTTAAATGCTTTGTCCTTCCTCCCCAG G AGA CGC AGC GCA GCC AAA CGC GCA GCC CCA
                                       g Arg Arg Ser Ala Ala Lys Arg Ala Ala Pro

                    30                  35                  40
GCC GCG AAG AAG GCT GCG AGT CGC CGT CGC CCA AAG AGT GCT AAG AAG/GTAGGTAATAAGATGT
Ala Ala Lys Lys Ala Ala Ser Arg Arg Arg Pro Lys Ser Ala Lys Lys
```

Exon 4

```
                                                            45                  50
AATTAAAAAAAAGTCCTGACAATATATTTTTCTCTTTTCAG GCT AAG CCC GCA GCA AGG AGG CGC AGC AGC
                                           Ala Lys Pro Ala Ala Arg Arg Arg Ser Ser

            55                  60                  65                  70
GTC AAA CCT AAA GCA GCA AAA GCA GCC ACC CAA GTC CGT CGC AGG AGC CGA CGA ATT CGC
Val Lys Pro Lys Ala Ala Lys Ala Ala Thr Gln Val Arg Arg Arg Ser Arg Arg Ile Arg

            75              +1                                                  +50
CGT GCG TCC GTG TCA AAG TAA TTCAATGGAAGACTGATCATTAAATCGTAACCCCTTCGAAAGATTAAACTTA
Arg Ala Ser Val Ser Lys  *
                                        +100
TCAAATTTCATTTTGTAGAACTGTCCAAATTTTCTAGAATATTGCAGAACTGAACATTTAAAACACATCCAAATTCGTAA
              +150                                                  +200
GCGAACAAGCAAGCAACGATGACCTACAATTTACAGTCGTTTCTTATTATTTCAAGTTTGCCTTTATTCAGTTTCAGTTT
                                    +250
CAGTTTATTTACTCTTTAATACCTCCTCGGAGGTGTCAGAGTCAAAACATACAATTAGATACACAGAAATATACAAAAAG
    +300                                                    +350
CAGCAAGATCAACAATAAACAAAAACAAAAAACAAAAATCATGCAAGCAAATCAGTACAATCAAAAAACTAACTTCAACC
                        +400                                                    +450
TAGTCACCAGACCCAATTGAAATAAATAAATCCAGAACAGAAATTGGATAGGCATGCATGAACTAGACGTAACACAAACA
                                        +500
TACGCACAACACAACGAGCAACCTACGTATGCACTAATGCCTAGCTATACACACTACATCATCAAATCAACTACGGCACT
              +550                                                  +600
CATTAACGATGATATACACTACAGAAGTGGCCAGGTTTCTTTGCACCTCCTTATTTCTAGACGTTAAAAGCTGGGCCATT
                                    +650
TTTAAAGTATTTGGACCCACCCTATAATAATGCTTCAAATAGTATTTCCGATCTGAGTTGAAAGCCTTACATTGAAAAAG
    +700                                                    +750
ATAATGAAATTACATCTCCTATCAGCCCACTAGTACAAAGTTCACAAAGTCTATCATTAAAGTCAATATTAAGAAATCGT
```

In turn, each singly-reacting fragment appeared to hybridize with the probe to a similar extent as revealed by the intensity of the autoradiographic signals estimated from densitometer tracings. Comparison of the level of hybridization of the genomic fragments with the quantified signal intensities of the cDNA standards yielded an average value of 0.92 copies of the ϕ_o gene per haploid genome. These results indicate that the *H. tubulosa* sperm ϕ_o protein is probably specified by a single copy gene.

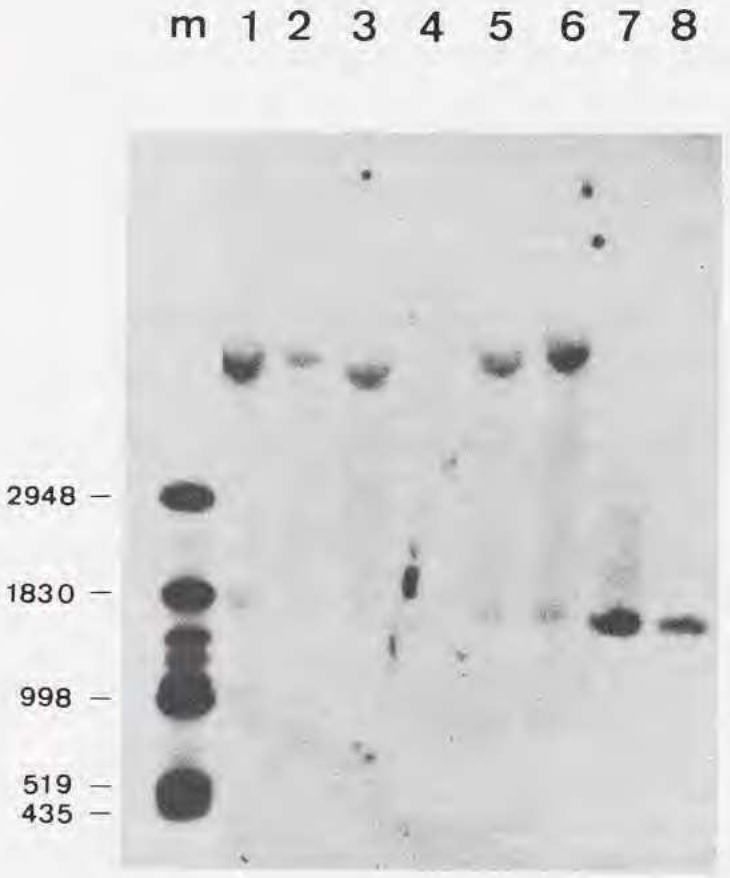

FIG. 3. — *H. tubulosa* ϕ_o gene copy number. Autoradiogram of Southern blots of sea cucumber sperm DNA restriction digests hybridized with *P*-probe of the cDNA (see text for details). Sperm DNA samples (20 µg) were digested with: lanes 1 to 6 *Bam*HI; *Hin*dIII; *Kpn*I; *Bam*HI + *Hin*dIII; *Bam*HI + *Kpn*I; *Hin*dIII + *Kpn*I, respectively. Graded amounts of the ϕ_o-cDNA cloned insert equivalent to 4 (lane 7) and 1 copies (lane 8) per haploid genome, supplemented with a mass excess of sheared calf thymus DNA were co-electrophoresed as hybridization standards. Sizes of restriction fragments from the ϕ_o-cDNA clone run in parallel as migration markers, are given in bp (lane m).

The structural arrangement and content of the ϕ_o gene closely coincide with the common features of a sizable number of post-meiotically expressed genes, typified by those encoding most histone-to-protamine transition proteins as well as protamine genes [20, 23]. Besides being expressed in a replication-uncoupled manner, these genes usually generate polyadenylated transcripts and most of them, although not all, are present as single copies containing coding regions often interrupted by intervening sequences. The overall organization of these genes becomes clearly divergent from that of the somatic histone genes. The functional implications, if any, of this divergence in gene arrangement remains unexplained, although potential correlations with chromatin transitions related to modulation and final arrest of gene activity during the spermiogenic process should be taken into account. Further studies are underway to characterize new genes encoding known sperm-specific protein variants.

ACKNOWLEDGEMENTS

This work has been supported by grants to L. C. from the EEC SCIENCE Programme (contract n°. SC1-CT91-0693) and the Spanish Dirección General de Investigación Científica y Técnica (PB91-0117)

REFERENCES

1. ALLAN, J., MITCHELL, T., HARBORNE, N., BOHM, L. & CRANE-ROBINSON, C., 1986. — Roles of H1 domains in determining higher order chromatin structure and H1 location. *Journal of Molecular Biology*, **187**: 591-601.
2. AUSIO, J. & SUBIRANA, J. A., 1982. — Conformational study and determination of the molecular weight of highly charged basic proteins by sedimentation equilibrium and electrophoresis. *Biochemistry*, **21**: 5910-5918.
3. AUSIO, J., 1986. — Structural variability and compositional homology of the protamine-like components of the sperm from the bivalve molluscs. *Comparative Biochemistry and Physiology*, **85B**: 439-449.
4. AUSIO, J. & VAN HOLDE, K. E., 1988. — The histones of the sperm from *Spisula solidissima* include a novel cysteine-containing H1 histone. *Cell Differentiation*, **23**: 175-190.
5. BENTON, W. D. & DAVIS, R. W., 1977. — Screening lambda-gt recombinant clones by hybridization to single plates *in situ*. *Science*, **196**: 180-182.
6. BIRNSTIEL, M. L., BUSSLINGER, M. & STRUB, K., 1985. — Transcription termination and 3' processing: The end is in site. *Cell*, **41**: 349-359.
7. BLOCH, D. P., 1976. — Histones of sperm. *In:* R. C. KING, *Handbook of Genetics*. New York, Plenum Press: 137-167.
8. BREATHNACH, R. & CHAMBON, P., 1981. — Organization and expression of eukaryotic split genes coding for proteins. *Annual Review of Biochemistry*, **50**: 349-383.
9. BRETZEL, G., 1972. — Über Thynnin, das Protamin des Thunfisches. *Hoppe-Seyler's- Zentralblatt Physiologische Chemie*, **353**: 933-943.
10. CASAS, M. T., MURA, C. V., SUBIRANA, J. A. & CORNUDELLA, L., 1989. — Purification and immunocytolocalization of protein ϕ_0 from sperm cells of the echinoderm *Holothuria tubulosa*. *Experimental Cell Research*, **182**: 14-25.
11. CHALLONER, P. B., MOSS, S. B. & GROUDINE, M., 1989. — Expression of replication-dependent histone genes in avian spermatids involves an alternate pathway of mRNA 3'-end formation. *Molecular and Cellular Biology*, **9**: 902-913.
12. CHILDS, G., MAXSON, R., COHN, R. & KEDES, L., 1981. — Orphons: dispersed genetic elements derived from tandem repetitive genes of eukaryotes. *Cell*, **23**: 651-663.
13. COLE, R. D., 1984. — A minireview of microheterogeneity in H1 histone and its possible significance. *Analytical Biochemistry*, **136**: 24-30.
14. COLES, L. S., ROBINS, A. J., MADLEY, L. K. & WELLS, J. R. E., 1987. — Characterization of the chicken histone H1 gene complement. *Journal of Biological Chemistry*, **262**: 9656-9663.
15. CORNUDELLA, L. & ROCHA, E., 1979. — Nucleosome organization during germ cell development in the sea cucumber *Holothuria tubulosa*. *Biochemistry*, **18**: 3724-3732.
16. DALTON, S., ROBINS, A. J., HARVEY, R. P. & WELLS, J. R. E., 1989. — Transcription from the intron-containing chicken histone $H2A_F$ gene is not S-phase regulated. *Nucleic Acids Research*, **17**: 1745-1756.
17. DIXON, G. H. & SMITH, M., 1968. — Nucleic acids and protamine in salmon testes. *Progress in Nucleic Acids Research*, **8**: 9-34.
18. DIXON, G. H., AIKEN, J. M., JANKOWSKI, J. M., MCKENZIE, D. I., MOIR, R. & STATES, C., 1986. — Organization and evolution of the protamine genes of salmonid fishes. *In:* G. REECK, G. A. GOODWIN & P. PUIGDOMENECH, *Chromosomal Proteins and Gene Expression*. New York, Plenum: 287-314.
19. DOENECKE, D. & TÖNJES, R., 1984. — Conserved dyad symmetry structures at the 3' end of the H5 histone genes. *Journal of Molecular Biology*, **178**: 121-135.
20. ERICKSON, R. P., 1990. — Post-meiotic gene expression. *Trends in Genetics*, **6**: 264-269.
21. FEINBERG, A. P. & VOGELSTEIN, B., 1983. — A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry*, **132**: 6-13.
22. FICHANT, G. A., 1992. — Constrains acting on the exon positions of the splice site sequences and local amino acid composition of the protein. *Human Molecular Genetics*, **1**: 259-267.
23. HECHT, N. B., 1990. — Regulation of haploid expressed genes in male germ cells. *Journal of Reproduction and Fertility*, **88**: 679-693.
24. JACKSON, I. J., 1991. — A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Research*, **19**: 3795-3798.

25. KISTLER, W. S., GEROCH, M. E. & WILLIAMS-ASHMAN, H. G., 1973. — Specific basic proteins from mammalian testes: isolation and properties of small, basic proteins from rat testes and epididymal spermatozoa. *Journal of Biological Chemistry*, **248**: 4532-4543.

26. LOENEN, W. A. M. & BLATTNER, F. R., 1983. — Lambda Charon vectors adapted for DNA cloning in recombination-deficient hosts. *Gene*, **26**: 171-179.

27. MAXSON, R., COHN, R., KEDES, L. & MOHUN, T., 1983. — Expression and organization of histone genes. *Annual Review in Genetics*, **17**: 239-277.

28. MOGENSEN, C., CARLOS, S. & AUSIO, J., 1991. — Microheterogeneity and interspecific variability of the nuclear sperm proteins from *Mytilus*. *FEBS Letters*, **282**: 273-276.

29. OLIVA, R. & DIXON, G. H., 1990. — Vertebrate protamine gene evolution. I. Sequence alignments and gene structure. *The Journal of Molecular Evolution*, **30**: 333-346.

30. OLIVA, R. & DIXON, G. H., 1991. — Vertebrate protamine genes and the histone-to-protamine replacement reaction. *Progress in Nucleic Acid Research and Molecular Biology*, **40**: 25-94.

31. PHELAN, J. J., SUBIRANA, J. A. & COLE, R. D., 1972. — An unusual group of lysine-rich histones from gonads of a sea cucumber, *Holothuria tubulosa*. *The European Journal of Biochemistry*, **31**: 63-68.

32. POCCIA, D., 1986. — Remodeling of nucleoproteins during gametogenesis, fertilization and early development. *International Review of Cytology*, **105**: 1-65.

33. PRATS, E., CORNUDELLA, L. & RUIZ-CARRILLO, A., 1989. — Nucleotide sequence of a cDNA for ϕ_0, a histone-to-protamine transition protein from sea cucumber spermatozoa. *Nucleic Acids Research*, **17**: 10097.

34. REED, K. C. & MANN, D. A., 1985. — Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Research*, **13**: 7207-7221.

35. RUIZ-CARRILLO, A., AFFOLTER, M. & RENAUD, J., 1983. — Genomic organization of the genes coding for the six main histones of the chicken: Complete sequence of the H5 gene. *Journal of Molecular Biology*, **170**: 843-859.

36. SAINZ, J., AZORIN, F. & CORNUDELLA, L., 1989. — Detection and molecular cloning of highly repeated DNA in the sea cucumber sperm. *Gene*, **80**: 57-64.

37. SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T., 1989. — *Molecular Cloning: A Laboratory Manual, 2nd Edition*. New York, Cold Spring Harbor.

38. SANGER, F., NICKLEN, S. & COULSON, A. R., 1977. — DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences-U.S.A.* **74**: 5463-5467.

39. SHIRAI, H., 1986. — Gonad-stimulating and maturation-inducing substance. *Methods in Cell Biology*, **27**: 73-88.

40. SUBIRANA, J. A., 1970. — Nuclear proteins from a somatic and a germinal tissue of the echinoderm *Holothuria tubulosa*. *Experimental Cell Research*, **63**: 253-260.

41. VON HOLT, C., 1986. — Histones in perspective. *BioEssays*, **3**: 120-124.

42. WOLFFE, A., 1992. — *Chromatin: Structure and Function*. London, Academic Press: 1-213.

43. ZWEIDLER, A., 1984. — Core histone variants of the mouse. *In:* G. S. STEIN, J. L. STEIN & W. F. MARZLUFF, *Histone Genes*. New York, Wiley & Sons: 339-372.